A New Approximate Min-Max Theorem with Applications in Cryptography

Maciej Skórski

maciej.skorski@mimuw.edu.pl
Cryptology and Data Security Group, University of Warsaw

Abstract. We propose a novel proof technique that can be applied to attack a broad class of problems in computational complexity, when switching the order of universal and existential quantifiers is helpful. Our approach combines the standard min-max theorem and convex approximation techniques, offering quantitative improvements over the standard way of using min-max theorems as well as more concise and elegant proofs.

Keywords: min-max theorems, convex approximation, leakage-resilient cryptography, hardness amplification

1 Introduction

1.1 The Min-Max Theorem.

The celebrate von Neumann min-max theorem [Neu28] states that every finite, two-player, zero-sum game has an equilibrium in mixed strategies. That is, the maximum value of the minimum expected gain for one player is equal to the minimum value of the maximum expected loss for the other. Any zero-sum game can be represented as a payoff matrix

$$A = [A(x,y)]_{x\in X, y\in Y}$$

where $A(x,y)$ is the payoff in case when the X-player chooses strategy $x \in X$ and the Y-player chooses strategy $y \in Y$, understood as a gain for the X-player and a loss of the Y-player. The basic moves $x \in X, y \in Y$ are called *pure strategies* (think of one of 3 options in the rock-paper-scissors game). We allow the players to use randomized strategies, which are called *mixed strategies* (think of picking a random answer in the rock-paper-scissors game) represented formally as distributions $p_X(\cdot), p_Y(\cdot)$ over X and Y respectively, and analyze the expected payoff

$$\mathbb{E}_{y\sim p_Y, x\sim p_X} A(x,y) = \sum_y \sum_x p_Y(y) p_X(x) A(x,y).$$

If the player X goes first, she can guarantee her gain to be at least

$$\mathsf{MaxGain}(X) = \max_{p_X} \min_{y} \mathbb{E}_{x \sim p_X} A(x, y),$$

and when the player Y goes first he guarantees his lost to be at most

$$\mathsf{MinLoss}(Y) = \min_{p_Y} \max_{x} \mathbb{E}_{y \sim p_Y} A(x, y),$$

where in both equations we used the fact that the second player always achieves the best response with some pure strategy. The min-max theorem guarantees that we have an equilibrium between the players.

Theorem (Min-Max Theorem [Neu28]). With the notation as above (and players using mixed strategies), we have

$$\mathsf{MaxGain}(X) = \mathsf{MinLoss}(Y).$$

Many more general versions of the min-max theorem exist. All of them assure the equality

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) = \inf_{y \in Y} \sup_{x \in X} f(x, y)$$

under certain conditions imposed on the sets X, Y (for example both convex and compact subsets of a locally convex topological space) and the function f (for example continuity, convexity in y and concavity in x). The proofs typically use fixed point theorems. Min-Max theorems have a lot of applications in game theory, statistical decision theory, economy and theoretical computer science. In this paper we focus on applications in cryptography, and the simplest version will be enough for our discussion.

1.2 Switching the order of quantifiers by the min-max theorem

The min-max theorem may be used to change the order of quantifiers (minimization corresponds to the existential quantifier and maximization corresponds to the universal quantifier). A very good example is the classical hardcore lemma due to Impagliazzo [Imp95]. The lemma stated informally says that if for every algorithm A there exists a large set of inputs on which A fails to compute a fixed function f, then in fact there exists a large set of inputs on which every algorithm fails to compute f with probability close to $\frac{1}{2}$. This particular lemma falls into a broad class of results in complexity theory which can be proven using the min-max theorem. We explain this technique before giving more examples.

The general framework. Let $\mathcal{A}$ be a class of test functions (for example poly-size circuits) over a set of possible inputs I and $\mathcal{C}$ be a class of distributions over I satisfying certain desired properties (for example samplability, high density, high entropy etc.), and v be a payoff function quantifies how well A performs on the input X (for example, unpredictability or distinguishing advantage). Suppose that we want to prove the existence of a distribution with certain properties for which every algorithm has bad (or alternatively good) performance.

Dream Statement. There is a distribution over inputs (with some certain properties) such that every algorithm performs badly/well.

$$\exists X \in \mathcal{C}\ \forall \mathrm{A} \in \mathcal{A} \quad v(\mathrm{A}, X) \leqslant c \tag{1}$$

In many cases, it is much easier to prove a weaker version, which gives the existence of a distribution with desired properties but only for a chosen algorithm.

Weak Statement. For every algorithm there is a distribution over inputs (with some certain properties) such that it performs badly/well.

$$\forall \mathrm{A} \in \mathcal{A}\ \exists X \in \mathcal{C}: \quad v(\mathrm{A}, X) \leqslant c \tag{2}$$

Note that this condition is considerably weaker. Indeed, we will see that in many applications proving the existence of a suitable distribution X for a fixed algorithm A is actually trivial. But the big question is whether Equation (2) implies Equation (1)

Does the Weak Statement imply the Dream Statement? Suppose that Equation (2) holds. Can we conclude that Equation (1) also holds, with possibly somewhat weaker class $\mathcal{A}$ and a weaker parameter c?

Note that we allow for some loss in quality (a weaker class of algorithms or a weaker payoff). Indeed, if both sets $\mathcal{C}$ and $\mathcal{A}$ are convex the answer is trivially "yes", by the min-max theorem. However, in most applications the set $\mathcal{A}$ consists of efficient algorithms (circuits of a bounded size) and is not convex, because taking a mixed strategy corresponds to combining many algorithms by (possibly) inefficient sampling. For the same reason, the set $\mathcal{C}$ might not be convex. However, we might "embed" non-convex sets $\mathcal{A}$ and $\mathcal{C}$ into "almost" convex hulls of $\mathcal{A}', \mathcal{C}'$ which are (hopefully) still sufficiently good for our purpose, by taking moderately long mixed strategies, instead of arbitrarily long. Indeed, let

$$\forall \mathrm{A} \in \mathcal{A}\ \forall X \in \mathcal{C}\ \exists \mathrm{A}' \in \operatorname{conv} \mathcal{A}'\ \exists X' \in \operatorname{conv} \mathcal{C}': \quad |v(A, X) - v(A', X')| \leqslant \delta \tag{3}$$

where the conv operator denotes the convex hull. We get the following

Approximate Min-Max Theorem If the condition (3) holds, then the Weak Statement implies the Dream Statement is true with $\mathcal{A}$ and $\mathcal{C}$ replaced by $\mathcal{A}'$ and $\mathcal{C}'$.

1.3 Our contribution

Summary. This framework is well known (cf. [BSW03,RTTV08,TTV08,Hol05,VZ] to mention only some papers closely related to our cryptographic applications). What we offer, is a *novel approximation technique*. Previous works used to find A' and X' in convex hulls by a trivial Chernoff approximation argument. We observe that much better results are obtained with a carefully chosen *convex*

approximation technique. Indeed, it turns out that in many cases the quantity $|v(A,X) - v(A',X')|$ can be upper bounded by the *Hölder Inequality* which involves moments of A and X. These moments may be better estimated based on properties of the sets $\mathcal{A}$ and $\mathcal{C}$ which leads to quantitative improvements. We stress that the key component is *the right choice of Hölder conjugates*, that is the exponents for the corresponding L_p, L_q spaces.

Advantages and Applications of our framework. Using our technique we prove a whole bunch of results, reproving what is already known in a more clear and concise way, improving quantitative bounds, or obtaining new results. Details are given in Section 2.

1.4 Related Works.

The work of [VZ] provides a tool to derive good bounds for certain sets $\mathcal{C}$, in the uniform settings. We stress that we consider only non-uniform adversaries here. In fact, our results can be probably made uniform by the use of constructive versions of auxiliary results on convex approximations we have applied (for example [DDGS97]). Anyway, uniform settings are not important for most of our applications like leakage-resilient crypto. While [VZ] gives hard bounds, we provide *a framework equipped with a different technique of handling* $\mathcal{C}$. Our technique can exploit *moment conditions*, which is impossible in [VZ]. We stress that the crucial component of our technique is the

2 Applications

We briefly recall some basic notation and conventions. We say that two distributions X_1, X_2 are (s,ϵ)-indistinguishable if for every A of size s we have $|\mathbb{E}\,\mathrm{A}(X_1) - \mathbb{E}\,\mathrm{A}(X_2)| \leqslant \epsilon$.

2.1 Impagliazzo Hardcore Lemma

Impagliazzo Hardcore Lemma. Suppose that are given a function $f : \{0,1\}^n \to \{0,1\}$ that is mildly hard to predict by a class of circuits; for every circuit A from our class, $\mathrm{A}(x)$ and $f(x)$ agree on at most, say, a 0.99 fraction of inputs x. This might happen when there is a set of noticeable size on which f is extremely hard to predict, meaning that there is (almost) no advantage over a random guess. This set could be as big as a $0.02 = 2(1-0.99)$ fraction of input. Indeed, if f cannot be guessed better than with probability $\frac{1}{2}$ on this set, then the probability that D agrees with f is at most $0.02 \cdot \frac{1}{2} + 0.98 \cdot 1 = 0.99$.

Quite surprisingly, this intuitive characterization is true. The first such result was proved by Impagliazzo [Imp95], with a sub-optimal hardcore density. An improved version with the optimal density of the hardcore set was found by Holenstein [Hol05]. Below we present the best possible result due to Klivans and Servedio, the lower bound was given in [LTW07].

Theorem 1 (Optimal Unpredictability Hardcore Lemma [KS03]). *Let $f : \{0,1\}^n \to \{0,1\}$ be ϵ-unpredictable by circuits of size s under a distribution V, that is*

$$\Pr_{x \leftarrow V}[\mathrm{A}(x) = f(x)] \leqslant 1 - \frac{\epsilon}{2}, \quad \text{for every A of size at most } s. \tag{4}$$

Then for any $\delta \in (0,1)$ there exists a event E of probability ϵ such that f is $1-\delta$ unpredictable under $V|E$ by circuits of size $s' = \Omega\left(s\delta^2/\log(1/\epsilon)\right)$, that is

$$\Pr_{x \leftarrow V|E}[\mathrm{A}(x) = f(x)] \leqslant \frac{1+\delta}{2}, \quad \text{for every A of size at most } s'. \tag{5}$$

Our Contribution. We reprove Theorem 1 using the framework discussed in Section 1.3. Our approach has the following advantages over the related works:

(a) It is derived from the *standard min-max theorem*. Previous proofs which achieve optimal parameters require involved iterative arguments [KS03,VZ].
(b) It is *modular and much simpler* than all alternative proofs. Indeed, the argument of Holenstein is non-optimal and involved. Also the argument given by Vadhan and Zheng depends on a non-trivial trick attributed to Nissan and Levy (which improves the hardcore density from $\frac{\epsilon}{2}$ to ϵ) and the machinery is much heavier. Our approach does not require this trick and follows the most intuitive strategy: show that there is a hardcore for every fixed adversary and then switch the order of quantifiers.
(c) We have identified *the reason for non-optimality in previous proofs*. Some authors even suggested that it might be impossible to get the tight parameters using the standard min-max theorem [VZ]. We show that this is not true. The problem is not with the standard min-max theorem but with an inadequate approximation argument in previous works, which do uniform approximation [Hol05].

A comparison is given below in Table 1.

Author	Technique	Hardcore Density	Complexity Loss
[Imp95]	boosting (constructive approx.)	$\Pr[E] = \frac{\epsilon}{2}$	$O(\delta^{-2} \cdot \mathrm{poly}(1/\epsilon))$
[Hol05]	standard min-max + Hardcore Optimization	$\Pr[E] = \epsilon$	$O(n\delta^{-2})$
[LTW07]	complicated boosting (constructive approx.)	$\Pr[E] = \epsilon$	$O(\log(1/\epsilon)\delta^{-2})$
[VZ]	complicated boosting (constructive approx.)	$\Pr[E] = \epsilon$	$O(\log(1/\epsilon)\delta^{-2})$
this paper	simple min-max + L_p-approx.	$\Pr[E] = \epsilon$	$O(\log(1/\epsilon)\delta^{-2})$

Table 1: Hardcore lemmas obtained by different techniques.

A sketch of proof. Assume without losing generality that $f : \{0,1\}^n \to \{-1,1\}$[1]. Define the payoff v as the unpredictability of f by A under X

$$v(A,X) \stackrel{\text{def}}{=} \Pr_{x \leftarrow X}[f(x) = A(x)] = \frac{1 + \mathbb{E}_{x \leftarrow X} \mathrm{A}(x) \cdot f(x)}{2},$$

[1] We consider $\{-1,1\}$ outputs for technical convenience. Equivalently we could state the problem for $\{0,1\}$.

and note that this definition makes sense also for circuits with real outputs. Let the property set $\mathcal{C}$ consists of conditional distributions of the form $X = V|E$ where $\Pr[E] \geqslant \epsilon$ and E may vary[2] ; note that $\mathcal{C}$ is convex. Define $\mathcal{A}$ as the set of real-valued[3] circuits of size s, and let $\mathcal{A}'$ be the set of circuits of size $s' = \frac{s}{\delta^{-2}\log(1/\epsilon)}$. It is not hard to see that the assumption (4) implies

Proposition 1 (Weak Statement). *For every* $\mathrm{A} \in \mathcal{A}$ *we have* $v(X, A) \leqslant 0$ *for some* $X \in \mathcal{C}$.

Now we analyze what happens when we replace $\mathcal{A}$ by $\mathrm{conv}(\mathcal{A}')$. We claim that

Proposition 2 (Approximation Step). *For every* $\mathrm{A}' \in \mathrm{conv}(\mathcal{A}')$ *we have* $v(X, \mathrm{A}') \leqslant \delta$ *for some* $X \in \mathcal{C}$.

To prove this, we show that the Hölder Inequality implies for A, A' and $X \in \mathcal{C}$

$$|v(X,\mathrm{A}) - v(X,\mathrm{A}')| \leqslant \frac{1}{2}\left(\mathbb{E}_{x\leftarrow V}\left(\frac{\mathbf{P}_{V|E}(x)}{\mathbf{P}_V(x)}\right)^q\right)^{\frac{1}{q}} \cdot \left(\mathbb{E}_{x\leftarrow V}\left|\mathrm{A}(x) - \mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}}$$

for any $p, q \geqslant 1$, $\frac{1}{p} + \frac{1}{q} = 1$. Now we can argue that

(a) $\left(\mathbb{E}_{x\leftarrow V}\left(\frac{\mathbf{P}_{V|E}(x)}{\mathbf{P}_V(x)}\right)^q\right)^{\frac{1}{q}} \leqslant \epsilon^{-\frac{1}{p}}$ (by the extreme points technique).

(b) $\left(\mathbb{E}_{x\leftarrow V}\left|\mathrm{A}(x) - \mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}} = O\left(\sqrt{\frac{p}{\ell}}\right)$ for some A which is of complexity ℓ relative to $\mathcal{A}'$[4] (by standard facts on convex-approximation [DHA97]).

Setting $\ell = \delta^{-2}\log(1/\epsilon)$ (so that $A \in \mathcal{A}$), taking $X = V|E$ which corresponds to A' according to Proposition 1, setting $p = 2\log(1/\epsilon)$ and putting this all together we get Proposition 5. This implies the following statement

Proposition 3 (Strong Statement). *For some* $X \in \mathcal{C}$ *we have* $v(X, A) \leqslant \delta$ *for every* $\mathrm{A} \in \mathcal{A}'$.

which proves Theorem 1 ($|v(X, A)| \leqslant \delta$ follows by considering $\mathcal{A}'$ closed under complements).

2.2 A (new) optimal hardcore lemma for metric pseudoentropy and applications to transformations.

Pseudoentropy notions extend classical information-theoretic entropy notions into computational settings. The following most widely used entropy notions capture what it means to be "computationally close" to a high entropy distribution.

[2] We can think of measures M such that $M(\cdot) \leqslant \mathbf{P}_V(\cdot)$ and $\sum_x M(x) \geqslant \epsilon$. Every $X \in \mathcal{C}$ can be written as $\mathbf{P}_X(\cdot) = M(\cdot)/\sum_x M(x)$ for one of these measures M.

[3] Following related works [FOR12,RTTV08] we use circuits with real outputs for technical reasons.

[4] That is, A is a convex combination of ℓ members of A'

Definition 1 (HILL Pseudoentropy [HILL99]). *Let Y be a distribution with the following property: there exists Y' of min-entropy at least k such that for every A of size at most s we have $|\mathbb{E}\,\mathrm{A}(Y) - \mathbb{E}\,Y'| \leqslant \epsilon$. Then we say that X has k bits of HILL entropy of quality (s,ϵ) and denote by $\mathbf{H}^{\mathrm{HILL}}_{s,\epsilon}(Y) \geqslant k$.*

Definition 2 (Metric Pseudoentropy [BSW03]). *Let Y be a distribution with the following property: for every A of size at most s there exists Y' of min-entropy at least k such that we have $|\mathbb{E}\,\mathrm{A}(Y) - \mathbb{E}\,Y'| \leqslant \epsilon$. Then we say that X has k bits of metric entropy of quality (s,ϵ) and denote by $\mathbf{H}^{\mathrm{Metric}}_{s,\epsilon}(Y) \geqslant k$.*

Pseudoentropy is an important research area, with applications in deterministic encryption, memory delegation [CKLR11], pseudorandom generators [HILL99,VZ]. Metric Pseudoentropy is much easier to deal with, and fortunately can be converted into HILL entropy with some loss in quality parameters (s,ϵ).

Our contribution. The following results shows that any distribution with metric pseudoentropy of 'moderate" quality has a kernel of HILL entropy with "strong" quality. We also conclude the optimal Metric-HILL transformation.

Theorem 2 (A HILL-pseudoentropy hardcore for metric pseudoentropy). *Suppose that $\mathbf{H}^{\mathrm{Metric}}_{s,\epsilon}(Y) \geqslant n - \Delta$, for some $Y \in \{0,1\}^n$. Then there is an event E, of probability $1-\epsilon$ such that $\mathbf{H}^{\mathrm{HILL}}_{s',\delta}(Y|E) \geqslant n - \Delta$ with $s' = \Omega(s\delta^2/(\Delta+1))$ for every δ. In particular, $\mathbf{H}^{\mathrm{HILL}}_{s',\epsilon+\delta}(Y) \geqslant n - \Delta$*

One possible application of this fact is amplifying hardness of pseudoentropy with poor quality. Imagine that we have many independent samples $X_1, X_2, \ldots, X_n$ from a distribution with a substantial entropy amount ($\Delta \ll n$) but of weak advantage $\epsilon = 0.99$. We can use the result above to show that pseudoentropy in $X_1, X_2, \ldots, X_n$ is roughly $(1-0.99)(n-\Delta)$ with good quality (see [Skob] for more details). Below we briefly compare this result with related works.

(a) Our result is *far stronger than the classical result due to Barak et al. [BSW03]* about the transformation. Not only we replace the factor n by Δ, but also show the existence of a hardcore in the intermediate step.
(b) This result *unifies and improves our recent results [Skoa,Skob].* The corollary $\mathbf{H}^{\mathrm{HILL}}_{s',\epsilon+\delta}(Y) \geqslant n - \Delta$ was the same (and optimal) but the hardcore E was found with worse complexity $s' = \Omega(s \cdot \delta^2/n)$.
(c) Our result *explains the nature of the Metric-HILL transformation.* The HILL pseudoentropy hardcore is an intermediate step in going from Metric pseudoentropy to HILL pseudoentropy.

Our result is illustrated in Figure 1. The parameters are optimal (see [Skob]).

A sketch of proof. Let $\mathcal{A}$ be the set of real-valued circuits of size s and let $\mathcal{A}'$ be the set of circuits of size $s' = s\delta^2/(\Delta+1)$. Let $\mathcal{C}$ consists of the conditional distributions X of the form $X'|E$, where $\Pr[E] \geqslant 1-\epsilon$ and $\mathbf{H}_\infty(X') \geqslant n - \Delta$; this set is convex. The payoff is defined as $v(X,A) \stackrel{def}{=} \mathbb{E}\mathrm{A}(Y) - \mathbb{E}\mathrm{A}(X)$. It is easy to see[5] that we have

[5] This is trivial for boolean A and somewhat more tricky for real-valued A. A short proof is given implicitly in [FOR12]

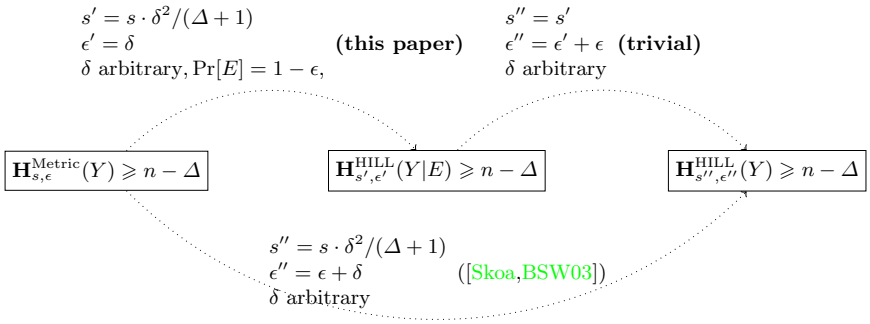

Fig. 1: The Metric-to-HILL pseudoentropy transformation.

Proposition 4 (Weak Statement). $\forall \mathrm{A} \in \mathcal{A}\ \exists X \in \mathcal{C} \quad v(X, A) \leqslant 0$.

Now we analyze what happens when $\mathcal{A}$ is replaced by $\mathrm{conv}(\mathcal{A}')$.

Proposition 5 (Approximation Step). *For every* $\mathrm{A}' \in \mathrm{conv}(\mathcal{A}')$ *we have* $v(X, \mathrm{A}') \leqslant \delta$ *for some* $X \in \mathcal{C}$.

To prove this, by the Hölder Inequality for any A, A' and $X \in \mathcal{C}$ we show

$$|v(X, \mathrm{A}) - v(X, \mathrm{A}')| \leqslant \left(\mathbb{E}_{x \leftarrow U} \left(2^n \mathbf{P}_{Y|E}(x)\right)^q\right)^{\frac{1}{q}} \cdot \left(\mathbb{E}_{x \leftarrow U} \left|\mathrm{A}(x) - \mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}}$$

for any $p, q \geqslant 1$, $\frac{1}{p} + \frac{1}{q} = 1$ and the uniform distribution U. Now we argue that

(a) $\left(\mathbb{E}_{x \leftarrow U} \left(2^n \mathbf{P}_{Y|E}(x)\right)^q\right)^{\frac{1}{q}} \leqslant 2^{\frac{\Delta}{p}}$ (by the extreme points technique).

(b) $\left(\mathbb{E}_{x \leftarrow V} \left|\mathrm{A}(x) - \mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}} = O\left(\sqrt{\frac{p}{\ell}}\right)$ for some A which is of complexity ℓ relative to $\mathcal{A}'$ (by standard facts on convex-approximation [DHA97]).

Setting $\ell = \delta^{-2}(\Delta+1)$ (so that $A \in \mathcal{A}$), taking $X = X'|E$ which corresponds to A' according to Proposition 1, setting $p = \Delta + 1$ and putting this all together we get Proposition 5. This implies the following statement

Proposition 6 (Strong Statement). $\exists X \in \mathcal{C}\ \forall \mathrm{A} \in \mathcal{A}' \quad v(X, A) \leqslant \delta$.

This directly implies Theorem 2 (as before, we consider A' closed under complements). More details can be found in Appendix D.

2.3 A (fixed) construction of a simulator for auxiliary inputs.

In [JP14] there is a theorem, which says that any short information Z about X can be efficiently simulated from X, Below we state the corrected version [Pie15].

Theorem 3 (Simulating auxiliary inputs, flaws fixed). *For any random variable* $X \in \{0,1\}^n$*, any correlated* $Z \in \{0,1\}^\lambda$ *and every choice of parameters* (ϵ, s) *there is a randomized function* $\mathsf{Sim} : \{0,1\}^n \to \{0,1\}^\lambda$ *of complexity* $O\left(s \cdot 2^{4\lambda}\epsilon^{-4}\right)$ *such that* Z *and* $\mathsf{Sim}(X)$ *are* (ϵ, s)*-indistinguishable given* X*.*

This result is the key component in the simplified analysis of the EUROCRYPT'09 stream cipher construction. Using Theorem 3, as described in [JP14], one proves the resilience of the cipher (assuming bounded leakage in every round) and if the underlying weak PRF is (s, ϵ)-secure against two queries on random inputs. The cipher security (s', ϵ') is related to (s, ϵ) by a polynomial loss in ϵ.

Our contribution. We describe a flaw in the proof and improve the corrected bound by a significant super polynomial factor. Below we briefly describe the significance of our result

(a) *Discovered flaws* in the recent (TCC'14) analysis of the EUROCRYPT'09 stream cipher. The alternative bounds seem correct but are much weaker. In particular, we get *no meaningful security with the AES* used as a weak PRF in this construction[6]. This raises the problem of *whether the cipher built on AES is secure or not*. We would need a simulator with a loss of only $O(\epsilon^{-2})$ not ϵ^{-4} in complexity.

(b) A *simpler* construction based on the min-max theorem. Based on the framework in Section 1.3 we give an alternative proof achieving the simulator complexity of $O\left(s \cdot 2^{2\lambda}\epsilon^{-4}\right)$. The gain of $2^{2\lambda}$ over the original approach, which is a power of ϵ for recommended values of parameters [JP14], comes from the use of convex approximation techniques. Our proof is considerably simpler and quantitatively better than in [JP14] (in particular we don't need to use the min-max theorem twice depending on what is the value of the game). Also, it is much simpler than the alternative approach of Vadhan and Zheng [VZ], yet yields comparable results for small leakages (see Table 2).

(c) A *clear bound on the security level*, in terms of the time-success ratio. We derive a clear formula which shows what fraction of the security of the original weak PRF is transformed into security of the stream cipher. This analysis shows that we are *far from good and provable secure* leakage-resilient stream ciphers as we lose over $\frac{5}{6}$ of original security. For more details, see Table 2.

In Table 2 we compare the strength of the simulator theorems in terms of implied security for this construction. To our knowledge, this is the first analysis of the time-success ratio for this technique. For more details we refer to Appendix E.

More on the flaws. In the claimed better bound $O\left(s \cdot 2^{3\lambda}\epsilon^{-2}\right)$ there is a mistake on page 18 (eprint version), when the authors enforce a signed measure to be a probability measure by a mass shifting argument. The number M defined there is in fact a function of x and is hard to compute, whereas the original

[6] The final bounds on the cipher security depends on the simulator complexity and are given by $\epsilon' = O\left(\sqrt{2^\lambda \epsilon}\right)$ and $s' = s \cdot 2^{-4\lambda}\epsilon'^4$. We can't prove then even very weak security $\epsilon' = 2^{-32}$ having $\lambda = 10$ bits of leakage!

Author	Technique	Simulator Complexity	Implied Security
[JP14]	Standard Min-Max + L_∞-approx.	$s_h = s \cdot 2^{4\lambda}\epsilon^{-4}$	$k' = \frac{k}{6} - \frac{5}{6}\lambda$
[VZ]	Complicated Boosting	$s_h = s \cdot 2^{\lambda}\epsilon^{-2} + 2^{\lambda}\epsilon^{-4}$	$k' = \frac{k}{6} - \frac{1}{3}\lambda$
this paper	Standard Min-Max + L_p-approx.	$s_h = s \cdot 2^{2\lambda}\epsilon^{-4}$	$k' = \frac{k}{6} - \frac{1}{2}\lambda$

Table 2: Security of the EUROCRYPT'09 stream cipher instantiated with a wPRF having 2^k keys and λ bits of leakage, obtained from different simulator results. Every attacker of size s succeeds with prob. at most $s/2^{k'}$

.

proof assumes that this is a constant independent of x. In the alternative bound $O\left(s \cdot 2^{3\lambda}\epsilon^{-2}\right)$ a fixable flaw is a missing factor of 2^λ in the complexity (page 16 in the eprint version), which is because what is constructed in the proof is only a probability mass function, not yet a sampler [Pie15].

A SKETCH OF THE PROOF. Let $\mathcal{A}$ be the set of real-valued circuits of size s and let $\mathcal{A}'$ be the set of circuits of size $s' = s \cdot 2^{-2\lambda}\epsilon^2$. Let $\mathcal{C}'$ consists of the distributions of the form $X, h(X)$, where h is computable in size $s \cdot 2^\lambda$; this set is *not* convex. Let $\mathcal{C}$ be the set of all circuits of size $s \cdot 2^{2\lambda}\epsilon^{-2}$. The payoff is defined as $v(h, A) \stackrel{def}{=} \mathbb{E}\mathrm{A}(X, h(X)) - \mathbb{E}\mathrm{A}(X, Z)$. It is easy to see that we have

Proposition 7 (Weak Statement). $\forall \mathrm{A} \in \mathcal{A}\ \exists h' \in \mathcal{C}' \quad v(h', A) \leqslant 0$.

Indeed, consider h^+_{A} which for every x outputs this value z for which $\mathrm{A}(x, z) = \max \mathrm{A}(x, \cdot)$ and h^-_{A} which for every x outputs this value z for which $\mathrm{A}(x, z) = \min \mathrm{A}(x, \cdot)$. Both are of complexity $O\left(2^\lambda\right)$. Since we have $\mathbb{E}\,\mathrm{A}(X, h^-(X)) \leqslant \mathbb{E}\,\mathrm{A}(X, Z)$ and $\mathbb{E}\,\mathrm{A}(X, Z) \leqslant \mathbb{E}\,\mathrm{A}(X, h^+(X))$, setting h' to be a distribution over h^+ and h^- that is $\Pr[h'(x) = z] = \theta \cdot \Pr[h^-(x) = z] + (1 - \theta) \cdot \Pr[h^+(x) = z]$, we get $v(h', \mathrm{A}) = 0$ with some θ. In the next step we replace $\mathcal{A}'$ by $\mathrm{conv}(\mathcal{A}')$.

Proposition 8 (Approximation 1). $\forall A \in \mathrm{conv}\mathcal{A}'\ \exists h' \in \mathcal{C}' : \quad v(h', \mathrm{A}') \leqslant \epsilon$.

This follows from the standard Chernoff Bound approximation argument[7] as

$$|v(h', A) - v(h', A')| = |\,\mathbb{E}\,\mathrm{A}(X, h'(X)) - \mathbb{E}\,\mathrm{A}'(X, h'(X))| \leqslant \sup_{x,z} |\mathrm{A}(x, z) - \mathrm{A}'(x, z)|.$$

Now we replace $\mathcal{C}'$ by conv $\mathcal{C}'$. Here a more delicate approximation is required.

Proposition 9 (Approximation 2). *For every A and every $h' \in$ conv $\mathcal{C}'$ there exists $h \in \mathcal{C}$ such that $v(h, \mathrm{A}) \leqslant v(h', \mathrm{A}) + \epsilon$.*

This follows because by the Hölder Inequality applied to $p = q = 2$ we obtain

$$|\,\mathbb{E}\,\mathrm{A}(X, h'(X)) - \mathbb{E}\,\mathrm{A}(X, h(X))| \leqslant 2^{\frac{\lambda}{2}} \cdot \left(\mathop{\mathbb{E}}_{x\sim X} \sum_z \left|\mathbf{P}_{x,h(x)}(x, z) - \mathbf{P}_{x,h'(x)(x,z)}\right|^2\right)^{\frac{1}{2}},$$

[7] A can be viewed as a distribution on $\mathcal{A}'$ we simply pick ℓ independent samples $\{\mathrm{A}_i\}_i$ and try to find an approximator of the form $\mathrm{A}' = \frac{1}{\ell}\sum_{i=1}^{\ell} \mathrm{A}_i$. It deviates by more than ϵ at (x, z) with probability $\exp(-2\ell\epsilon^2)$. We combine this with the union bound.

and by the standard results on convex approximation [DDGS97] the second factor is at most $\ell^{-\frac{1}{2}}$ for some h of complexity ℓ with respect to $\mathcal{C}'$. We put $\ell = 2^{\lambda}\epsilon^{-2}$. From the proven propositions we obtain the final result.

Proposition 10 (Strong Statement). $\exists h \in \mathcal{C}\ \forall \mathrm{A} \in \mathcal{A}' \quad v(h, \mathrm{A}) \leqslant 2\epsilon$.

2.4 More Applications

For more applications we refer an interested reader to Appendix A. They include the optimal Dense Model Theorem, a better auxiliary input simulator for bounded-variance adversaries (new), and a proof that every high-conditional entropy source can be efficiently simulated (new, extending [TTV08]).

References

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy*, RANDOM, 2003, pp. 200–215.

CKLR11. K.-M. Chung, Y. Y. Kalai, F.-H. Liu, and R. Raz, *Memory delegation*, CRYPTO'11, 2011.

DDGS97. M.J. Donahue, C. Darken, L. Gurvits, and E. Sontag, *Rates of convex approximation in non-hilbert spaces*, Constructive Approximation **13** (1997).

DHA97. D. Docampo, D. R. Hush, and C. T. Abdallah, *Intelligent methods in signal processing and communications*, Birkhauser Boston Inc., 1997.

DP08. S. Dziembowski and K. Pietrzak, *Leakage-resilient cryptography*, FOCS '08, 2008.

FOR12. B. Fuller, A. O'Neill, and L. Reyzin, *A unified approach to deterministic encryption (...)*, TCC'12, 2012.

HILL99. J. Hastad, R. Impagliazzo, L. A. Levin, and M. Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. (1999).

Hol05. T. Holenstein, *Key agreement from weak bit agreement*, STOC '05, 2005.

Imp95. R. Impagliazzo, *Hard-core distributions for somewhat hard problems*, FOCS 36, 1995.

JP14. D Jetchev and K. Pietrzak, *How to fake auxiliary input*, TCC 2014, 2014.

KS03. Adam R. Klivans and Rocco A. Servedio, *Boosting and hard-core set construction*, Mach. Learn. **51** (2003), no. 3, 217–238.

LTW07. Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu, *On the complexity of hard-core set constructions*, ICALP'07, Springer-Verlag, 2007, pp. 183–194.

Neu28. John Neumann, *Zur Theorie der Gesellschaftsspiele*, 295–320+.

Pie15. Krzysztof Pietrzak, *private communication, may*, 2015.

RTTV08. O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan, *Dense subsets of pseudorandom sets*, FOCS '08, 2008.

Skoa. Maciej Skorski, *Metric pseudoentropy: Characterizations, transformations and applications*, ICITS 2015.

Skob. ———, *Nonuniform indistinguishability and unpredictability hardcore lemmas: New proofs and applications to pseudoentropy*, ICITS 2015.

TTV08. L. Trevisan, M. Tulsiani, and S. Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, CCC '09, 2008.

VZ. Salil Vadhan and ColinJia Zheng, *A uniform min-max theorem with applications in cryptography*, CRYPTO 2013.

Zha11. Jiapeng Zhang, *On the query complexity for showing dense model*, ECCC (2011).

A More Applications

.

A.1 Dense Model Theorem

Given a pair of two distributions W and V over the same finite domain we say that W is δ-dense in V if and only if $\Pr[W = x] \leqslant \Pr[V = x]/\delta$[8]. The efficient version of the famous dense model theorem specialized to the boolean case, can be formulated as follows:

Theorem 4 (Dense Model Theorem.). *Let $\mathcal{D}'$ be a class of n-bit boolean functions, R be uniform over $\{0,1\}^n$, X be an n-bit random variable and let X' be δ-dense in X. If X and R are $(\mathcal{D}, \epsilon)$-indistinguishable then there exists a distribution R' which is δ-dense in R such that X' and R' are $(\mathcal{D}', \epsilon')$-indistinguishable, where $\epsilon' = (\epsilon/\delta)^{O(1)}$ and $\mathcal{D}$ consists of all functions of the form $g(D_1, \ldots, D_\ell)$ where $D_i \in \mathcal{D}'$, $\ell = \mathrm{poly}(1/\delta, 1/\epsilon)$ and g is some function.*

Using our framework we can reprove the Dense Model Theorem with optimal parameters due to Zhang [Zha11]. The proof is very similar to the one in Theorem 2 so we omit the details; we note that we need the Hölder Inequality with $p = 2\log(1/\delta)$. A similar technique appears in [Skoa], though we do note use Metric pseudoentropy here.

Corollary 1. *Dense Model Theorem (Theorem 4) holds with $\epsilon' = O(\epsilon/\delta)$, g being a linear threshold and $\ell = O(\log(1/\delta)/(\epsilon/\delta)^2$.*

Author	Technique	Function g	ℓ as complexity of $\mathcal{D}'$ w.r.t $\mathcal{D}$	ϵ' vs ϵ
Tao and Ziegler	Complicated	Inefficient	$\ell = \mathrm{poly}(1/(\epsilon/\delta), \log(1/\delta))$	$\epsilon' = O(\epsilon/\delta)$
[RTTV08]	Min-Max Theorem	Linear threshold	$\mathrm{poly}(1/(\epsilon/\delta), \log(1/\delta))$	$\epsilon' = O(\epsilon/\delta)$
[FOR12], [DP08]	Metric Entropy	Linear threshold	$\ell = O(n/(\epsilon/\delta)^2)$	$\epsilon' = O(\epsilon/\delta)$
[Zha11]	Boosting	Linear threshold	$\ell = O(\log(1/\delta)/(\epsilon/\delta)^2$	$\epsilon' = O(\epsilon/\delta)$
This paper	Standard Min-Max + L_p-approx	Linear threshold	$\ell = O(\log(1/\delta)/(\epsilon/\delta)^2$	$\epsilon' = O(\epsilon/\delta)$

Table 3: Different versions of the Dense Model Theorem

A.2 Simulating auxiliary inputs against bounded-variance distinguishers

An interesting result is obtained when working more carefully in the proof of Theorem 3. Namely, imposing additional restriction on the second moment of test functions we obtain a refined bound

[8] The term "δ-dense" comes from the fact that V can be written as a convex combination of W with weight δ and some other distribution with weight $1 - \delta$

Theorem 5 (Simulating auxiliary inputs against bounded-variance distinguishers). *For any random variable $X \in \{0,1\}^n$, any correlated $Z \in \{0,1\}^\lambda$, any class $\mathcal{A}$ be of functions $\mathrm{A} : \{0,1\}^n \times \{0,1\}^m \to [0,1]$ such that $\forall \mathrm{A} \in \mathcal{A} : \mathbb{E}_X \operatorname{Var}\mathrm{A}(x,U) \leqslant \sigma^2$, and every ϵ there is a randomized function $\mathsf{Sim} : \{0,1\}^n \to \{0,1\}^\lambda$ of complexity $O\left(s \cdot 2^{4\lambda}\sigma\epsilon^{-4}\right)$ relative to $\mathcal{A}$ such that Z and $\mathsf{Sim}(X)$ are ϵ-indistinguishable given X by functions $\mathcal{A}$.*

This result is interesting in the context of recent improvements in key derivation, so called square security, where the second moment condition is widely used.

A.3 Simulating High Entropy Distribution with Auxiliary Information

From our framework we derive the following result, which is the extension of the theorem in [TTV08] into a conditional case (in the presence of auxiliary information). We stress that this result *cannot* be derived from the techniques used in [TTV08], because this approach will not preserve the same marginal distribution Z, when applied in the conditional setting.

Theorem 6 (High conditional min-entropy is simulatable). *Let $X \in \{0,1\}^n$ and $Z \in \{0,1\}^m$ be correlated random variables and $\mathbf{H}_\infty(X|Z) = n - \Delta$. Then there exists a distribution Y, Z such that*

(a) There is a circuit Sim of complexity $O\left(n(n+m)2^{2\Delta}\epsilon^{-5}\right)$ and such that $\mathsf{Sim}(Z) \stackrel{d}{=} Y$
(b) (X,Z) and (Y,Z) are (s,ϵ)-indistinguishable
(c) We have $\mathbf{H}_\infty(Y|Z) \geqslant n - \Delta - 6$.

Here we show only how to simulate given a fixed distinguisher. The rest of the proof follows by the use of a convex approximation argument (with $p = q = 2$) and allows us to save a factor of $2^{2\Delta}$ in the simulator complexity.

Proof (Proof of the weak statement). Let $\Delta = n - k$. By replacing ϵ with 2ϵ we can assume that $\mathcal{D} = \sum_{i=1}^{j} \alpha_i \mathcal{D}_i$ where $\alpha_i = 1 - (i-1)\epsilon$ for $i = 1, \ldots, \lceil 1/\epsilon \rceil$ and $\mathcal{D}_i$ are boolean such that $\mathbf{1} = \sum_i \mathcal{D}_i$. Define

$$d(i) = \Pr[\mathrm{D}(U) \geqslant \alpha_i]. \tag{6}$$

and let M be the smallest number i such that $d(i) \geqslant 2^{-\Delta}$. Note that if we didn't care about computational efficiency then the best answer would be

$$Y^+ \stackrel{d}{=} \frac{d(M-1)}{2^{-\Delta}} \cdot U_{\mathcal{D}_1 + \ldots + \mathcal{D}_{M-1}} + \frac{2^{-\Delta} - d(M-1)}{2^{-\Delta}} \cdot U_{\mathcal{D}_M} \tag{7}$$

because then

$$\begin{aligned} \mathbf{E}\mathrm{D}(Y^+) &= \frac{\sum_{i=1}^{M-1} \alpha_i |\mathrm{D}_i| + \left(2^k - \sum_{i=1}^{M-1} \alpha_i |\mathrm{D}_i|\right) \alpha_M}{2^k} \\ &= \max_{Y:\ \mathbf{H}_\infty(Y) \geqslant k} \mathbf{E}\mathrm{D}(Y) \end{aligned} \tag{8}$$

The approach we chose is quite obvious - we efficiently approximate the distribution Y^+. For any i, sample $x_1, \ldots, x_\ell$ where $\ell > 2^{\Delta} n \log(1/\epsilon)/\epsilon$ and let

$$\tilde{d}(i) = \ell^{-1} \sum_{j=1}^{\ell} \mathbf{1}_{\{\mathrm{D}(x_i) \geqslant \alpha_i\}} \tag{9}$$

Now let M' be the smallest number such that $\tilde{d}(M') > \frac{3}{4} \cdot 2^{-\Delta}$. Note that that M' is well defined with probability $1 - 2^{-n}$, and then we have

$$\tilde{d}(M'-1) < \frac{3}{4} \cdot 2^{-\Delta} < \tilde{d}(M') \tag{10}$$

Now we define Y as follows:

$$Y \stackrel{d}{=} \begin{cases} \frac{\tilde{d}(M'-1)}{2^{-\Delta}} \cdot \widetilde{U}_{D_1+\ldots+D_{M'-1}} + \left(1 - \frac{\tilde{d}(M'-1)}{2^{-\Delta}}\right) \cdot \widetilde{U}_{\mathcal{D}_{M'}}, & 2^{-\Delta}\epsilon < \tilde{d}(M'-1) < 2^{-\Delta}/16 \\ \widetilde{U}_{D_1+\ldots+D_{M'-1}}, & 2^{-\Delta}/16 < \tilde{d}(M'-1) \\ \widetilde{U}_{\mathcal{D}_{M'}}, & 2^{-\Delta}\epsilon > \tilde{d}(M'-1) \end{cases} \tag{11}$$

Observe that if $d(i) < 2^{-\Delta}/4$ then with probability $1-2^{-n}$ we get $\tilde{d}(i) < 2^{-\Delta}/2$. Thus, the probability that $d(M') < 2^{-\Delta}/4$ and $\tilde{d}(M') > 2^{-\Delta}/2$ is at most $2^{-n}\log(1/\epsilon)$ and we can assume that $d(M') > 2^{-\Delta}/4$. Similarly, if $d(i) > 2^{-\Delta}$ then with probability $1 - 2^{-n}$ we have $\tilde{d}(i) > \frac{3}{4} \cdot 2^{-\Delta}$ which means $M' \leqslant i$. Therefore, with probability $1-2^{-n}\log(1/\epsilon)$ we can assume that $d(M'-1) < 2^{-\Delta}$. Now we split the analysis into the following cases

(a) $\tilde{d}(M'-1) < 2^{-\Delta}\epsilon$ and $d(M'-1) < 2 \cdot 2^{-\Delta}\epsilon$. Since $|\mathrm{D}_{M'}| = 2^n(d(M') - d(M'-1)) \geqslant 2^{n-\Delta}/8$, we see that $U_{\mathcal{D}_{M'}}$ is samplable in time $\mathcal{O}\left(2^{\Delta}\log(1/\epsilon)\right)$ and that $\mathbf{H}_\infty(U_{\mathcal{D}_{M'}}) \geqslant k-3$. Note that

$$\begin{aligned} \mathbf{E}\mathrm{D}(Y^+) &= \mathbf{E}\mathrm{D}(Y^+)\mathbf{1}_{\mathrm{D}(Y^+) \geqslant \alpha_{M'-1}} + \mathbf{E}\mathrm{D}(Y^+)\mathbf{1}_{\mathrm{D}(Y^+) \leqslant \alpha_{M'}} \\ &\leqslant 2\epsilon + \alpha_{M'} \\ &\leqslant 3\epsilon + \mathbf{E}\mathrm{D}(Y) \end{aligned} \tag{12}$$

(b) $\tilde{d}(M'-1) > 2^{-\Delta}/16$ and $2^{-\Delta} > d(M'-1) > 2^{-\Delta}/32$. Then we have $|\mathrm{D}_1| + \ldots + |\mathrm{D}_{M'-1}| \geqslant 2^{n-\Delta-5}$ and thus $\mathbf{H}_\infty(\widetilde{U}_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}) \geqslant n - \Delta - 5$ and $\widetilde{U}_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}$ is samplable in time $\mathcal{O}\left(2^{\Delta}\log(1/\epsilon)\right)$. Since $|\mathrm{D}_1| + \ldots + |\mathrm{D}_{M'-1}| \leqslant 2^{n-\Delta}$, we have

$$\begin{aligned} \mathbf{E}\mathrm{D}(Y^+) &\leqslant \mathbf{E}\mathrm{D}(U_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}) \\ &\leqslant \mathbf{E}\mathrm{D}(\widetilde{U}_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}) + \epsilon \end{aligned} \tag{13}$$

(c) $2^{-\Delta}\epsilon < \tilde{d}(M'-1) < 2^{-\Delta}/16$ and $2^{-\Delta}\epsilon/2 < d(M'-1) < 2^{-\Delta}/8$ and $\tilde{d}(M'-1) \leqslant 2d(M'-1)$. We have $|\mathrm{D}_1| + \ldots + |\mathrm{D}_{M'-1}| = 2^n d(M'-1) > 2^{n-\Delta}\epsilon/2$ and $|D_{M'}| = 2^n(d(M') - d(M'-1)) \geqslant 2^{n-\Delta}/8$, therefore Y is samplable in

time $\mathcal{O}\left(2^{\Delta}\log(1/\epsilon)/\epsilon\right)$. Moreover, we have $\mathbf{H}_\infty(\widetilde{U}_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}) \geqslant \log(|\mathrm{D}_1|+\ldots+|\mathrm{D}_{M'-1}|)$ and $\mathbf{H}_\infty(\widetilde{U}_{\mathcal{D}_{M'}}) \geqslant \log|D_{M'}|$. Hence $\mathbf{H}_\infty(\widetilde{U}_{\mathcal{D}_1+\ldots+\mathcal{D}_{M'-1}}) \geqslant n+\log d(M'-1)$ and $\mathbf{H}_\infty(\widetilde{U}_{\mathcal{D}_{M'}}) \geqslant n-\Delta-3$ and

$$\begin{aligned}\Pr[Y=x] &\leqslant \frac{\tilde{d}(M'-1)}{d(M'-1)}\cdot 2^{-n+\Delta}+2^{-n+\Delta+3}\\ &\leqslant 2^{-n+\Delta+4}\end{aligned} \tag{14}$$

Suppose now that $d(M'-1) < 2^{-\Delta}\epsilon/2$. Then, by the Chernoff Bound with probability $1-2^{-n}$ we have $\tilde{d}(M'-1) < 2^{-\Delta}\epsilon/2 + d(M'-1) < 2^{-\Delta}\epsilon$ and we are in case (a). If $2^{-\Delta}\epsilon/2 < d(M'-1) < 2^{-\Delta}/32$ then with probability $1-2^{-n}$ we have $\frac{1}{2} < \frac{\tilde{d}(M'-1)}{d(M'-1)} < 2$ and it is easy to check that we can be either in (a) or in (c), depending on $\tilde{d}(M'-1)$. If $2^{-\Delta}/32 < d(M'-1) < 2^{-\Delta}/8$ then with probability $1-2^{-n}$ we are either in (c) or in (b). If $2^{-\Delta}/8 < d(M'-1) < 2^{-\Delta}$ then with probability $1-2^{-n}$ we can be only in (b).

B Convex Approximation Rates

We use the following fact on convex approximation rates.

Lemma 1 (Convex approximation in L^p spaces [DDGS97]). *Let $\mathcal{X}$ be a finite domain, ν be a distribution on $\mathcal{X}$. Fix a number $1 \leqslant p < +\infty$ and for any function f on $\mathcal{X}$ define $\|f\|_p = (\mathbb{E}_{x\leftarrow\nu}|f(x)|^p)^{\frac{1}{p}}$. Let $\mathcal{G}$ be any set of real functions on $\mathcal{X}$, let $\overline{g}$ be a convex combinations of functions from $\mathcal{G}$ and $K>0$ be such that for all $g\in\mathcal{G}$ we have $\|\overline{g}-g\|_p \leqslant K$. Then for any $\ell>0$ there exists a convex combination $g'=\sum_{i=1}^{\ell}\alpha_i g_i$ of functions $g_1,\ldots,g_k\in\mathcal{G}$ such that*

$$\|\overline{g}-g'\|_p \leqslant \frac{KC_p}{\ell^{1-\frac{1}{t}}}$$

where $t=\min(2,p)$ and $C_p=1$ if $1\leqslant p\leqslant 2$, $C_p=\sqrt{2}[\Gamma((p+1)/2)/\sqrt{\pi}]^{1/p}$ for $2<p<+\infty$.

C Proof of Theorem 2

To finish the proof it remains to justify the estimates

$$\left(\mathbb{E}_{x\leftarrow U}\left(2^n\mathbf{P}_{Y|E}(x)\right)^q\right)^{\frac{1}{q}} \leqslant 2^{\frac{\Delta}{p}} \tag{15}$$

and

$$\left(\mathbb{E}_{x\leftarrow V}\left|\mathrm{A}(x)-\mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}} = O\left(\sqrt{\frac{p}{\ell}}\right) \text{ for A of complexity } \ell \text{ r. t. } \mathcal{A}' \tag{16}$$

The first follows by noticing that the quantity is convex with respect to $Y|E\in\mathcal{C}$. Thus, the maximum is attained at one of extreme points which is, in this case, a flat distribution. The second fact follows from Lemma 1.

D Proof of Theorem 2

To finish the proof it remains to justify the estimates

$$\left(\mathbb{E}_{x\leftarrow U}\left(2^n \mathbf{P}_{Y|E}(x)\right)^q\right)^{\frac{1}{q}} \leqslant 2^{\frac{\Delta}{p}} \tag{17}$$

and

$$\left(\mathbb{E}_{x\leftarrow V}\left|\mathrm{A}(x)-\mathrm{A}'(x)\right|^p\right)^{\frac{1}{p}} = O\left(\sqrt{\frac{p}{\ell}}\right) \text{ for A of complexity } \ell \text{ r. t. } \mathcal{A}' \tag{18}$$

The first follows by noticing that the quantity is convex with respect to $Y|E \in \mathcal{C}$. Thus, the maximum is attained at one of extreme points which is, in this case, a flat distribution. The second fact follows from Lemma 1.

E Time-Success Ratio for Auxiliary Input Simulator Analysis of Stream Ciphers

E.1 Preliminaries

Weak pseudorandom functions, are *indistinguishable* from random functions, when queried on random inputs and fed with iniform secret key.

Definition 3 (Weak pseudorandom functions). *A function* $\mathrm{F}:\{0,1\}^k\times\{0,1\}^n\to\{0,1\}^m$ *is an* (ϵ,s,q)*-secure weak PRF if its outputs on* q *random inputs are indistinguishable from random by any distinguisher of size* s*, that is*

$$\left|\Pr\left[\mathrm{D}\left((X_i)_{i=1}^q,\mathrm{F}((K,X_i)_{i=1}^q)=1\right]-\Pr\left[\mathrm{D}\left((X_i)_{i=1}^q,(R_i)_{i=1}^q\right)=1\right]\right|\leqslant\epsilon$$

where the probability is over the choice of the random $X_i\leftarrow\{0,1\}^n$*, the choice of a random key* $K\leftarrow\{0,1\}^k$ *and* $R_i\leftarrow\{0,1\}^m$ *conditioned on* $R_i=R_j$ *if* $X_i=X_j$ *for some* $j<i$*.*

Stream ciphers generate a keystream in a recursive manner. The security means that the output stream should be indistinguishable from uniform[9].

Definition 4 (Stream ciphers). *A* stream-cipher $\mathsf{SC}:\{0,1\}^k\to\{0,1\}^k\times\{0,1\}^n$ *is a function that need to be initialized with a secret state* $S_0\in\{0,1\}^k$ *and produces a sequence of output blocks* $X_1,X_2,\ldots$ *computed as*

$$(S_i,X_i):=\mathsf{SC}(S_{i-1}).$$

A stream cipher SC *is* (ϵ,s,q)*-secure if for all* $1\leqslant i\leqslant q$*, the random variable* X_i *is* (s,ϵ)*-pseudorandom given* $X_1,\ldots,X_{i-1}$ *(the probability is also over the choice of the initial random key* S_0*).*

[9] We note that in a more standard notion the entire stream $X_1,\ldots,X_q$ is indistinguishable from random. This is implied by the notion above by a standard hybrid argument, with a loss of a multiplicative factor of q in the distinguishing advantage.

Now we define the security of leakage resilient stream ciphers, which follow the "only computation leaks" assumption.

Definition 5 (Leakage-resilient stream ciphers). *A leakage-resilient stream-cipher is $(\epsilon, s, q, \lambda)$-secure if it is (ϵ, s, q)-secure as defined above, but where the distinguisher in the j-th round gets λ bits of arbitrary deceptively chosen leakage about the secret state accessed during this round. More precisely, before $(S_j, X_j) := \mathsf{SC}(S_{j1})$ is computed, the distinguisher can choose any leakage function f_j with range $\{0,1\}^\lambda$, and then not only get X_j, but also $\Lambda_j := f_j(\hat{S}_{j1})$, where $\hat{S}_{j1}$ denotes the part of the secret state that was modified (i.e., read and/or overwritten) in the computation $\mathsf{SC}(S_{j1})$.*

Finally, we recall the standard notion of time-success ratio. It is very useful in quantifying how much security is transformed from the underlying primitive to the constructed object by the reduction.

Definition 6 (Time-Success Ratio). *We say that a cryptographic protocol has k bits of security (or that it is 2^k-secure) if for every s and any adversary* A *of size s the advantage* A *(probability of winning in the security game) is at most $\epsilon \leqslant s/2^k$.*

E.2 Time-Success Ratio Analysis

Suppose that we have a simulator which guarantees if we have a simulator with complexity $t_h = O\left(t \cdot A\epsilon^{-\alpha} + B\epsilon^{-\beta}\right)$ then, according to [JP14], we have a (s', ϵ', q)-secure stream cipher where

$$\epsilon' = O\left(q \cdot \sqrt{2^\lambda \epsilon}\right), \quad s' = \Omega\left(s \cdot A^{-1}(\epsilon')^\alpha\right) - A^{-1}B(\epsilon')^{\alpha-\beta} \tag{19}$$

Suppose that we want to prove $2^{k'}$-security in the sense of Definition 6. That is, we need to prove $s'/\epsilon' \geqslant 2^{k'}$ for every time-advantage pair (s', ϵ') such that $s' \geqslant 1$, where k' is possibly big. Note that for a weak PRF we can assume the security $s \approx 2^k\epsilon$ for every ϵ, that is that the best attack is by a brute-force search over the key space (see [JP14] for more justifiction). One can argue that, under the transformation (19), the worst-case adversary profile is when $\epsilon' \approx 2^{-k'}$ and $s' \approx 1$. Pugging this in Equation (19), and using the fact that $s' \geqslant 1$ we obtain

$$2^k \cdot 2^{-2k'-\lambda} > A \cdot \left(2^{-k'}\right)^{-\alpha} + B \cdot \left(2^{-k'}\right)^{-\beta}.$$

Substituting different values of A, B, α, β which correspond to the particular bounds, we get the values in Table 2.